Agency for Healthcare Research and Quality
Advancing Excellence in Health Care

NATIONAL QUALITY MEASURES CLEARINGHOUSE

General

Title

Diagnostic imaging: percentage of final reports for screening mammograms that are classified "probably benign."

Source(s)

American College of Radiology (ACR), American Medical Association-convened Physician Consortium for Performance ImprovementÂ® (PCPIÂ®), National Committee for Quality Assurance (NCQA). Diagnostic imaging performance measurement set. Reston (VA): American College of Radiology (ACR); 2015 Feb. 58 p. [89 references]

Measure Domain

Primary Measure Domain

Clinical Quality Measures: Process

Secondary Measure Domain

Does not apply to this measure

Brief Abstract

Description

This measure is used to assess the percentage of final reports for screening mammograms that are classified as "probably benign."

Rationale

The "probably benign" assessment category is reserved for findings that have a high probability (greater than or equal to 98%) chance of being benign and should not be used as a category for indeterminate findings. Inappropriate designation of findings as "probably benign" can result in unnecessary follow-up of lesions that could have been quickly classified or delayed diagnosis and treatment of cancerous lesions (Kerlikowske et al., 2005). Published guidance documents (American College of Radiology [ACR], 2013; D'Orsi et al., 2003) emphasize the need to conduct a complete diagnostic imaging evaluation before making a probably benign (Category 3) assessment; making it inadvisable to use the probably benign

categorization when interpreting a screening mammogram. Immediate completion of a diagnostic imaging evaluation for abnormal screening mammograms eliminates potential anxiety that women would endure with the short interval follow-up that is recommended for "probably benign" findings.

The following clinical recommendation statements are quoted verbatim from the referenced clinical guidelines and represent the evidence base for the measure:

A category 3, 4, or 5 assessment is not recommended for a screening mammogram, even though in some instances a highly suspicious abnormality may be identified that will warrant a recommendation for biopsy. Rather, all patients with screening abnormalities should be given a BI-RADS® category 0 assessment and recalled for further diagnostic studies (ACR, 2013).

All the previously cited studies emphasize the need to conduct a complete diagnostic imaging evaluation before making a probably benign (category 3) assessment; hence it is recommended not to render such an assessment in interpreting a screening mammography examination. The practice of rendering category 3 assessments directly from screening examination also has been shown to result in adverse outcomes: 1) unnecessary follow-up of many lesions that could have been promptly assessed as benign, and 2) delayed diagnosis of a small number of cancers that otherwise may have been smaller in size and less likely to be advanced in stage (ACR, 2013; D'Orsi et al., 2003).

The use of assessment category 3, probably benign, has been clarified in the lexicon of the 2013 edition. It is emphasized that this is *not* an indeterminate category used simply when the radiologist is unsure whether the render a benign (BI-RADS® category 2) or suspicious (BI-RADS® category 4) assessment, but one that is reserved for specific imaging findings known to have a greater than essentially 0% but less than or equal to 2% likelihood of representing malignancy (ACR, 2013; D'Orsi et al., 2003).

For mammography, there is robust literature describing three findings (noncalcified circumscribed solid mass, focal asymmetry and solitary group of punctate calcifications) that have likelihoods of malignancy in the defined (less than or equal to 2%) probably benign range, for which short interval (6-month) follow-up mammography and then periodic mammographic surveillance represents appropriate management (ACR, 2013; D'Orsi et al., 2003; Kerlikowske et al., 2005; Baum et al., 2011). Use of assessment category 3 for mammographic findings other than these three should be considered only if the radiologist has personal experience to justify a watchful-waiting approach, preferably involving observation of a sufficient number of cases of an additional mammographic finding to suggest a likelihood of malignancy within the defined (less than or equal to 2%) probably-benign range. Two large-scale studies performed in the United States have validated that in the usual-care setting, category 3 assessments indeed are associated with a likelihood of malignancy of less than or equal to 2% (ACR, 2013; D'Orsi et al., 2003).

Evidence for Rationale

American College of Radiology (ACR), American Medical Association-convened Physician Consortium for Performance ImprovementÂ® (PCPIÂ®), National Committee for Quality Assurance (NCQA). Diagnostic imaging performance measurement set. Reston (VA): American College of Radiology (ACR); 2015 Feb. 58 p. [89 references]

American College of Radiology (ACR). ACR practice guideline for the performance of screening and diagnostic mammography. Reston (VA): American College of Radiology (ACR); 2013 Oct. 11 p.

Baum JK, Hanna LG, Acharyya S, Mahoney MC, Conant EF, Bassett LW, Pisano ED. Use of BI-RADS 3-probably benign category in the American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial. Radiology. 2011 Jul;260(1):61-7. PubMed

Dâ€™Orsi CJ, Bassett LW, Berg WA, et al. BI-RADS: mammography. In: Dâ€™Orsi CJ, Mendelson EB, Ikeda DM, et al, editor(s). Breast Imaging Reporting and Data System: ACR BI-RADS â€“ Breast

Imaging Atlas. 4th ed. Reston (VA): American College of Radiology (ACR); 2003.

Kerlikowske K, Smith-Bindman R, Abraham LA, Lehman CD, Yankaskas BC, Ballard-Barbash R, Barlow WE, Voeks JH, Geller BM, Carney PA, Sickles EA. Breast cancer yield for screening mammographic examinations with recommendation for short-interval follow-up. Radiology. 2005 Mar;234(3):684-92. PubMed

Primary Health Components

Screening mammography; "probably benign" assessment category

Denominator Description

All final reports for screening mammograms

Numerator Description

Final reports classified as "probably benign" (see the related "Numerator Inclusions/Exclusions" field)

Evidence Supporting the Measure

Type of Evidence Supporting the Criterion of Quality for the Measure

A clinical practice guideline or other peer-reviewed synthesis of the clinical research evidence

A formal consensus procedure, involving experts in relevant clinical, methodological, public health and organizational sciences

Additional Information Supporting Need for the Measure

Importance of Topic
As imaging technology continues to advance, the United States healthcare system has seen an increase in both the type and frequency of imaging studies being performed. The increase in utilization of imaging studies is accompanied by a corresponding increase in cost and exposure to radiation for both patients and healthcare professionals.

- From 1980 to 2006, the number of radiologic procedures performed in the United States showed a ten-fold increase while the annual per-capita effective dose from radiologic and nuclear medicine procedures increased by 600% (Mettler et al., 2009).
- From 1996 to 2010, the number of computerized tomographic (CT) examinations tripled, while the number of ultrasounds nearly doubled (Smith-Bindman et al., 2012).
- From 1996 to 2010, advanced diagnostic imaging (i.e., CT, magnetic resonance imaging [MRI], nuclear medicine, and ultrasound) accounted for approximately 35% of all imaging studies (Smith-Bindman et al., 2012).
- From 1980 to 2006, the proportion of radiation exposure that is attributable to medical sources increased from 17% to 53% (Mettler et al., 2009).
- In 2006, while CT scans only accounted for approximately 17% of all radiologic procedures performed in the United States, they accounted for over 65% of the total effective radiation dose from radiologic procedures (Mettler et al., 2009).
- In 2006, the estimated per-capita effective radiation dose for radiologic procedures in the United States was nearly 20% higher than the average for other well-developed countries (Mettler et al., 2009).

Diagnostic imaging was prioritized as a topic area for measure development due to a high level of utilization, rising costs, and the need for measures to help promote appropriate use of imaging and improve outcomes.

Opportunity for Improvement
Although a mammogram assessment category of "probably benign" is not recommended for use in interpreting screening mammograms, it is associated with approximately 2.3% of screening mammograms (Baum et al., 2011). Additionally, compliance is shown to be poor among women referred to short-interval follow-up (Baum et al., 2011).

Evidence for Additional Information Supporting Need for the Measure

American College of Radiology (ACR), American Medical Association-convened Physician Consortium for Performance ImprovementÂ® (PCPIÂ®), National Committee for Quality Assurance (NCQA). Diagnostic imaging performance measurement set. Reston (VA): American College of Radiology (ACR); 2015 Feb. 58 p. [89 references]

Baum JK, Hanna LG, Acharyya S, Mahoney MC, Conant EF, Bassett LW, Pisano ED. Use of BI-RADS 3-probably benign category in the American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial. Radiology. 2011 Jul;260(1):61-7. PubMed

Mettler FA, Bhargavan M, Faulkner K, Gilley DB, Gray JE, Ibbott GS, Lipoti JA, Mahesh M, McCrohan JL, Stabin MG, Thomadsen BR, Yoshizumi TT. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources--1950-2007. Radiology. 2009 Nov;253(2):520-31. PubMed

Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, Greenlee RT, Kruger RL, Hornbrook MC, Roblin D, Solberg LI, Vanneman N, Weinmann S, Williams AE. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. JAMA. 2012 Jun 13;307(22):2400-9. PubMed

Extent of Measure Testing

The American Medical Association (AMA)-convened Physician Consortium for Performance Improvement (PCPI) collaborated on a measure testing project in 2011 with Telligen to ensure four radiology measures were reliable and evaluated for accuracy of the measure numerator, denominator, and exception case identification. The testing project was conducted utilizing electronic health record data and claims data. Inter-rater reliability was tested. Three sites in three states participated in the testing of the measures. All three sites were in urban settings. Site A was a group practice with 10 physicians. Site B was a hospital-based group practice with 90 physicians. Site C was a hospital-based practice with greater than 1000 physicians.

Reliability Testing
The purpose of reliability testing was to evaluate whether the measure definitions and specifications, as prepared by the PCPI, yield stable, consistent measures. Data abstracted from chart records were used to calculate inter-rater reliability for the measures.

Some of the measures in this set are being made available without any prior testing. The PCPI recognizes the importance of testing all of its measures and encourages testing of the diagnostic imaging measurement set for feasibility and reliability by organizations or individuals positioned to do so. The *Measure Testing Protocol for PCPI Measures* was approved by the PCPI in 2010 and is available on the PCPI Web site (see Position Papers at www.physicianconsortium.org); interested parties are encouraged to review this document and to contact PCPI staff. The PCPI will welcome any opportunity to promote the initial testing of these measures and to ensure that any results available from

testing are used to refine the measures before implementation.

Evidence for Extent of Measure Testing

American College of Radiology (ACR), American Medical Association-convened Physician Consortium for Performance ImprovementÂ® (PCPIÂ®), National Committee for Quality Assurance (NCQA). Diagnostic imaging performance measurement set. Reston (VA): American College of Radiology (ACR); 2015 Feb. 58 p. [89 references]

State of Use of the Measure

State of Use

Current routine use

Current Use

not defined yet

Application of the Measure in its Current Use

Measurement Setting

Ambulatory/Office-based Care

Ambulatory Procedure/Imaging Center

Hospital Inpatient

Hospital Outpatient

Long-term Care Facilities - Other

Skilled Nursing Facilities/Nursing Homes

Professionals Involved in Delivery of Health Services

not defined yet

Least Aggregated Level of Services Delivery Addressed

Individual Clinicians or Public Health Professionals

Statement of Acceptable Minimum Sample Size

Does not apply to this measure

Target Population Age

Unspecified

Target Population Gender

Either male or female

National Strategy for Quality Improvement in Health Care

National Quality Strategy Aim

Better Care

National Quality Strategy Priority

Health and Well-being of Communities
Prevention and Treatment of Leading Causes of Mortality

Institute of Medicine (IOM) National Health Care Quality Report Categories

IOM Care Need

Staying Healthy

IOM Domain

Effectiveness

Data Collection for the Measure

Case Finding Period

Unspecified

Denominator Sampling Frame

Patients associated with provider

Denominator (Index) Event or Characteristic

Diagnostic Evaluation

Denominator Time Window

not defined yet

Denominator Inclusions/Exclusions

Inclusions
All final reports for screening mammograms

Exclusions
Unspecified

Exceptions
None

Exclusions/Exceptions

not defined yet

Numerator Inclusions/Exclusions

Inclusions
Final reports classified as "probably benign"

Probably Benign Classification: Mammography Quality Standards Act (MQSA) assessment category of "probably benign"; Breast Imaging Reporting and Data System (BI-RADS®) category 3; or Food and Drug Administration (FDA)-approved equivalent assessment category.

Exclusions
Unspecified

Numerator Search Strategy

Fixed time period or point in time

Data Source

Administrative clinical data

Electronic health/medical record

Imaging data

Paper medical record

Registry data

Type of Health State

Does not apply to this measure

Instruments Used and/or Associated with the Measure

Unspecified

Computation of the Measure

Measure Specifies Disaggregation

Does not apply to this measure

Scoring

Rate/Proportion

Interpretation of Score

Desired value is a lower score

Allowance for Patient or Population Factors

not defined yet

Standard of Comparison

not defined yet

Identifying Information

Original Title

Measure #2: inappropriate use of "probably benign" assessment category in screening mammograms.

Measure Collection Name

Diagnostic Imaging Performance Measurement Set

Submitter

American College of Radiology - Medical Specialty Society

Developer

American College of Radiology - Medical Specialty Society

National Committee for Quality Assurance - Health Care Accreditation Organization

Physician Consortium for Performance ImprovementÂ® - Clinical Specialty Collaboration

Funding Source(s)

Unspecified

Composition of the Group that Developed the Measure

Diagnostic Imaging Measure Development Work Group Members

William Golden, MD (*Co-chair*) (internal medicine)
David Seidenwurm (*Co-chair*) (diagnostic radiology)
Michael Bettmann, MD
Dorothy Bulas, MD (pediatric radiology)
Rubin I. Cohen, MD, FACP, FCCP, FCCM
Richard T. Griffey, MD, MPH (emergency medicine)
Eric J. Hohenwalter, MD (vascular interventional radiology)
Deborah Levine, MD, FACR (radiology/ultrasound)
Mark Morasch, MD (vascular surgery)
Paul Nagy, MD, PhD (radiology)
Mark R. Needham, MD, MBA (family medicine)
Hoang D. Nguyen (diagnostic radiology/payer representative)
Charles J. Prestigiacomo, MD, FACS (neurosurgery)
William G. Preston, MD, FAAN (neurology)
Robert Pyatt, Jr., MD (diagnostic radiology)
Robert Rosenberg, MD (diagnostic radiology)
David A. Rubin, MD (diagnostic radiology)
B Winfred (B.W.) Ruffner, MD, FACP (medical oncology)
Frank Rybicki, MD, PhD, FAHA (diagnostic radiology)
Cheryl A. Sadow, MD (radiology)
John Schneider, MD, PhD (internal medicine)
Gary Schultz, DC, DACR (chiropractic)
Paul R. Sierzenski, MD, RDMS (emergency medicine)
Michael Wasylik, MD (orthopedic surgery)

Diagnostic Imaging Measure Development Work Group Staff

American College of Radiology: Judy Burleson, MHSA; Alicia Blakey, MS

American Medical Association-convened Physician Consortium for Performance Improvement: Mark Antman, DDS, MBA; Kathleen Blake, MD, MPH; Kendra Hanley, MS; Toni Kaye, MPH; Marjorie Rallins, DPM; Kimberly Smuk, RHIA; Samantha Tierney, MPH; Stavros Tsipas, MA

National Committee for Quality Assurance: Mary Barton, MD

Financial Disclosures/Other Potential Conflicts of Interest

None of the members of the Diagnostic Imaging Work Group had any disqualifying material interest under the Physician Consortium for Performance Improvement (PCPI) Conflict of Interest Policy.

Endorser

National Quality Forum - None

NQF Number

not defined yet

Date of Endorsement

2014 Dec 23

Measure Initiative(s)

Physician Quality Reporting System

Adaptation

This measure was not adapted from another source.

Date of Most Current Version in NQMC

2015 Feb

Measure Maintenance

This measure is reviewed and updated every 3 years.

Date of Next Anticipated Revision

2018

Measure Status

This is the current release of the measure.

This measure updates a previous version: American College of Radiology, Physician Consortium for Performance Improvement®, National Committee for Quality Assurance. Radiology physician performance measurement set. Chicago (IL): American Medical Association (AMA); 2010 Sep. 45 p.

The measure developer reaffirmed the currency of this measure in March 2017.

Measure Availability

Source available from the American College of Radiology (ACR) Web site .

For more information, contact ACR at 1891 Preston White Drive, Reston, VA 20191; Phone: 703-648-8900; E-mail: info@acr.org; Web site: www.acr.org .

NQMC Status

This NQMC summary was completed by ECRI Institute on February 1, 2008. The information was verified by the measure developer on April 10, 2008.

This NQMC summary was updated by ECRI Institute on April 23, 2009. The information was verified by the measure developer on September 16, 2009.

This NQMC summary was retrofitted into the new template on June 10, 2011.

This NQMC summary was edited by ECRI Institute on April 27, 2012.

Stewardship for this measure was transferred from the PCPI to the ACR. ACR informed NQMC that this

measure was updated. This NQMC summary was updated again by ECRI Institute on October 13, 2015. The information was verified by the measure developer on November 19, 2015.

The information was reaffirmed by the measure developer on March 3, 3017.

Copyright Statement

This NQMC summary is based on the original measure, which is subject to the measure developer's copyright restrictions.

Production

Source(s)

American College of Radiology (ACR), American Medical Association-convened Physician Consortium for Performance ImprovementÂ® (PCPIÂ®), National Committee for Quality Assurance (NCQA). Diagnostic imaging performance measurement set. Reston (VA): American College of Radiology (ACR); 2015 Feb. 58 p. [89 references]

Disclaimer

NQMC Disclaimer

The National Quality Measures Clearinghouseâ„¢ (NQMC) does not develop, produce, approve, or endorse the measures represented on this site.

All measures summarized by NQMC and hosted on our site are produced under the auspices of medical specialty societies, relevant professional associations, public and private organizations, other government agencies, health care organizations or plans, individuals, and similar entities.

Measures represented on the NQMC Web site are submitted by measure developers, and are screened solely to determine that they meet the NQMC Inclusion Criteria.

NQMC, AHRQ, and its contractor ECRI Institute make no warranties concerning the content or its reliability and/or validity of the quality measures and related materials represented on this site. Moreover, the views and opinions of developers or authors of measures represented on this site do not necessarily state or reflect those of NQMC, AHRQ, or its contractor, ECRI Institute, and inclusion or hosting of measures in NQMC may not be used for advertising or commercial endorsement purposes.

Readers with questions regarding measure content are directed to contact the measure developer.